

Wrap-up

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Linear Regression vs Deep Neural Networks

- 비선형 데이터의 관계 또는 함수에 대해서 근사(approximation) 가능
 - 신경망의 깊이, 너비에 따라 capacity가 결정
 - 더 깊고 넓은 네트워크를 통해 더 복잡한 함수를 근사할 수 있음
 - 하지만 파라미터가 늘어남에 따른 최적화가 어려워질 것
- 여전히 같은 방법(gradient descent)을 통해 최적화
 - 하지만 DNN은 non-convex한 loss surface를 가짐

Backpropagation

- 편미분의 chain-rule을 통해, 합성함수의 미분을 나누어 접근 가능
 - 이를 통해 효율적인 미분 계산이 가능
- Gradient Vanishing
 - Sigmoid와 TanH는 기울기가 항상 1보다 작거나 같음
 - 따라서 backprop. 과정에서 반복될 경우, 기울기 값은 작아질 것
 - ReLU의 활용을 통해 어느정도 해결할 수 있음